

**IMPROVING THE EFFICIENCY OF
HISTOGRAM OF ORIENTED GRADIENT
FEATURE FOR HUMAN DETECTION**

LAI CHI QIN

UNIVERSITI SAINS MALAYSIA

2016

**IMPROVING THE EFFICIENCY OF HISTOGRAM OF
ORIENTED GRADIENT FEATURE FOR HUMAN DETECTION**

by

LAI CHI QIN

Thesis submitted in fulfillment of requirements

for the degree of

Master of Science

September 2016

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank everyone who have helped me and supported me throughout this project.

I would like to express my gratitude and appreciation to my Master Project supervisor Dr. Teoh Soo Siang for his great effort in guiding and helping me in this research. He is a great supervisor and guides me towards the correct path of this research with his well-organized project guideline. His professional suggestions and guidance helped me to overcome problem faced during my progress of the project. In addition, I would also like to thanks my co-supervisor Dr. Dzati Athiar Ramli for providing useful advice and guidance on my research.

Next, I would like to thank the Ministry of Education Malaysia and Universiti Sains Malaysia for their financial support under the FRGS Research Grant 203/PELECT/6071292. Lastly, I would like to express my special appreciation and gratitude to my dear family and friends. They had given me unconditional support throughout my project.

TABLES OF CONTENTS

	Page
ACKNOWLEDGEMENTS	ii
TABLES OF CONTENTS	iii
LIST OF TABLES	vii
LIST OF FIGURES	viii
LIST OF ABBREVIATIONS	x
ABSTRAK	xi
ABSTRACT	xiii
 CHAPTER ONE : INTRODUCTION	
1.1 Overview	1
1.2 Problem Statement	3
1.3 Research Objective	3
1.4 Project Scope	4
1.5 Thesis Outline	4
 CHAPTER TWO : LITERATURE REVIEW	
2.1 Introduction	6
2.2 Feature extraction for human detection	7
2.2.1 Histogram of Oriented Gradient	7
2.2.2 Improvements on the HOG features	12

2.2.2(a)	Hardware acceleration	12
2.2.2(b)	Combination of HOG feature extraction with other method	13
2.2.2(c)	Reducing the complexity of the algorithm	17
2.2.2(d)	Summary of the improvements	17
2.2.3	Local Binary Pattern (LBP)	18
2.3	SVM classification	20
2.4	Feature reduction using PCA	23
2.5	Summary	24

CHAPTER THREE: METHODOLOGY

3.1	Introduction	26
3.2	The proposed feature extraction method	27
3.2.1	Image pre-processing	28
3.2.2	Gradient computation	28
3.2.3	Dividing the input image into cells and blocks	29
3.2.4	Construct the histogram of the gradients' orientation	30
3.2.5	Using selective number of histogram bins for different regions in the image	31
3.2.6	Block normalization	32
3.4	Feature selection by Principal Component Analysis	35
3.5	SVM Classification	35

3.6	Summary	37
-----	---------	----

CHAPTER FOUR: RESULTS AND DISCUSSION

4.1	Overview	38
4.2	Experiment Set up	39
4.2.1	Dataset	39
4.2.2	Performance Matrices	40
4.2.3	SVM classifier parameters selection and training	42
4.3	Experiments, results and discussion	44
4.3.1	Experiment to evaluate the performance of feature extracted using different number of histogram bins	45
4.3.2	Experiment to evaluate different ways of histogram normalization	49
4.3.3	Experiment to evaluate the performance of features extracted using selective number of histogram bins for different regions in the image	55
4.3.4	Experiment to evaluate feature reduction using PCA	58
4.3.5	Comparison of the proposed method with other existing methods for pedestrian detection	61
4.3.6	Testing of the proposed method on road scene images	64
4.4	Summary	67

CHAPTER FIVE: CONCLUSION

5.1	Conclusion	68
-----	------------	----

5.2	Future work	69
	References	70

LIST OF TABLES

		Page
Table 2.1	Summary of approach by researchers	18
Table 4.1	Miss Rate at FPPW = 10^{-3} (%) for features extracted using different number of histogram bins	49
Table 4.2	Processing time of features extracted using different number of histogram bins	49
Table 4.3	Missed rate at FPPW = 10^{-3} for different normalization methods tested on features extracted using different number of orientation bins.	53
Table 4.4	Processing time of different normalization methods for features extracted using different number of orientation bins	54
Table 4.5	Miss Rate at FPPW = 10^{-3} for different combinations of (higher/lower) number of histogram bins	56
Table 4.6	Processing time of different combination of regional bins	56
Table 4.7	Miss Rate at FPPW = 10^{-3} (%) for different number of feature subsets	60
Table 4.8	Processing time for different number of feature subset	60
Table 4.9	Miss Rate at FPPW = 10^{-3} for the proposed method, original HOG, LBP and the integral image HOG method	62
Table 4.10	Processing time for the proposed method, original HOG, LBP and the integral HOG method	62

LIST OF FIGURES

	Page
Figure 2.1	Summary of HOG extraction 8
Figure 2.2	Block formed from 4 neighboring cells 10
Figure 2.3	The 9 histograms bins and their respective ranges of orientation angles 10
Figure 2.4	Basic concept of LBP feature 19
Figure 2.5	Multi-scale LBP. $R = 2$, $P = 4$ 20
Figure 2.6	An example of the optimal separating hyperplane and margin for a two-dimensional feature space 22
Figure 3.1	Overall Design Flow Chart 27
Figure 3.2	Dividing the gradient into cells and overlapping blocks (note: not all blocks are shown in the figure) 29
Figure 3.3	Dividing 8 orientation angles into 4 histogram bins 30
Figure 3.4	(a) From the average image, the important regions in the image that may contain human features are determined. Blocks located in these regions (as shown in the shaded blocks in (b)) are extracted with higher number of histogram bins while the rest are extracted with lower number of histogram bins 32
Figure 3.5	Grouping of blocks to perform normalization 34
Figure 3.6	Summary of the proposed method 34
Figure 4.1	Example of positive training samples 39
Figure 4.2	Example of negative training samples 40
Figure 4.3	An example of DET curve 41

Figure 4.4	Plots to visualize the C and γ parameters search. (a) First stage - coarse search. (b) Second stage - finer search	43
Figure 4.5	Division of orientation angles into different bins (a) 4 bins 8 angles (b) 8 bins 16 angles (c) 16 bins 32 angles (d) 32 bins 64 angles	46
Figure 4.6	Effect of feature extraction using different number of histogram bins on the detection performance	48
Figure 4.7	The grouping of blocks for histogram normalization	50
Figure 4.8	The effect of normalization methods on the classification performance. Tested on features extracted using different number of histogram bins (a) 4-bin (b) 8-bin (c) 16-bin (d) 32-bin	51
Figure 4.9	Performance of features extracted using different combinations of higher/lower number of histogram bins	56
Figure 4.10	Percentage of variance explained by each PC	59
Figure 4.11	Performance of different number of feature subset	59
Figure 4.12	Comparison of the proposed method with the original HOG, LBP and the integral image HOG method	62
Figure 4.13	Movement of the sliding detection window	64
Figure 4.14	Human Detection in real road scene	65

LIST OF ABBREVIATIONS

AdaBoost	Adaptive Boosting
CPU	Central Processing Unit
CUDA	Compute Unified Device Architecture
DET	Detection Error Trade-off Graph
FPGA	Field-Programmable Gate Array
FPPW	False Positive Per Window
GPU	Graphics Processing Unit
HDTV	High Definition Video
HOG	Histogram of Oriented Gradient
LBP	Local Binary Pattern
OpenCV	Open Source Computer Vision
PCA	Principal Component Analysis
PCs	Principal Components
RBF	Radial Basis Function
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine
VGA	Video Graphics Array
ViBE	Visual Background Extractor
DWT	Discrete Wavelet Transform

MENINGKATKAN KECEKAPAN CIRI HISTOGRAM KECERUNAN TERHALA UNTUK PENGESANAN MANUSIA

ABSTRAK

Histogram Kecerunan Terhala (HOG) yang asalnya dicadangkan oleh Dalal dan Triggs telah digunakan dengan meluas dalam aplikasi pengesanan manusia berasaskan penglihatan. Walau bagaimanapun, kaedah tersebut menghasilkan himpunan ciri yang besar dan memerlukan pengiraan yang intensif serta memakan masa. Oleh itu, kaedah ini tidak sesuai untuk digunakan dalam aplikasi-aplikasi masa nyata. Penyelidikan ini mencadangkan satu kaedah baru yang boleh mengurangkan masa pengekstrakan ciri HOG tanpa menjejaskan terlalu banyak prestasi pengesanan. Kaedah yang dicadangkan ialah menjalankan pengekstrakan ciri dengan menggunakan bilangan bin histogram terpilih. Bilangan bin histogram yang lebih tinggi yang mampu mengekstrak ciri-ciri yang mengandungi lebih banyak maklumat orientasi telah digunakan untuk pengekstrakan ciri di kawasan imej yang berkemungkinan mengandungi manusia, manakala pengekstrakan ciri-ciri di kawasan yang lain menggunakan bilangan bin histogram yang lebih rendah. Cara ini akan mengurangkan saiz ciri tanpa menjejaskan terlalu banyak prestasi pengesanan. Kemudian, Analisis Komponen Utama (PCA) akan digunakan untuk menyusun dan memilih ciri-ciri yang boleh mewakili keseluruhan set ciri. Pengelas linear Sokongan Vektor Mesin (SVM) telah digunakan untuk menilai prestasi kaedah yang dicadangkan dalam penyelidikan ini. Eksperimen telah dijalankan menggunakan data set manusia INRIA. Keputusan eksperimen menunjukkan bahawa kaedah yang dicadangkan ini mampu mengurangkan masa pengekstrakan ciri sebanyak 2.6 kali

ganda berbanding dengan kaedah HOG yang asal manakala pengurangan sebanyak 7 kali ganda berbanding dengan kaedah LBP dan pengurangan sebanyak 2.5 kali ganda berbanding dengan kaedah HOG imej integral. Pada masa yang sama, kaedah yang dicadangkan ini memberikan prestasi pengesanan yang setanding.

IMPROVING THE EFFICIENCY OF HISTOGRAM OF ORIENTED GRADIENT FEATURE FOR HUMAN DETECTION

ABSTRACT

Histogram of Oriented Gradient (HOG) feature which was originally proposed by Dalal and Triggs is widely used in vision-based human detection. However, HOG feature extraction method produced a large feature pool which is computationally intensive and very time consuming, causing it not so suitable for real time application. This research proposed a method to reduce the HOG feature extraction time without affecting too much on its detection performance. The proposed method performs feature extraction using selective number of histogram bins. Higher number of histogram bins which can extract more detailed orientation information is applied on the regions of image that may contain human figure. The rest of the regions in the image are extracted using lower number of histogram bins. This will reduce the feature size without compromising too much on the performance. To further reduce the feature size, Principal Component Analysis (PCA) is used to rank the features and select only the representative features. A linear Support Vector Machine (SVM) classifier is used to evaluate the performance of the proposed method. Experiment was conducted using the INRIA human dataset. The test results show that the proposed method is able to reduce the feature extraction time by 2.6 times compared to the original HOG, 7 times compared to the LBP method and 2.5 times faster than the integral image HOG while providing comparable detection performance.

CHAPTER ONE

INTRODUCTION

1.1 Overview

According to the Malaysian traffic police statistic [1], there are a total of 562 pedestrians killed annually in urban areas in 2012 to 2015 in Malaysia. 40% of the casualties were children, while pedestrians aged from 66 to 70 years old made up the highest number of fatalities. A driver assistance system that can detect pedestrians and warn the driver of any potential collision will be able to reduce the number of accidents involving pedestrians.

On the other hand, the statistic from the U.S. Department of State [2] shows that there were 1122 terrorist attacks, 2727 deaths and 2899 injuries per month worldwide in 2014. To prevent these tragedies, it has become necessary to install surveillance cameras and its associated system which can detect abnormal human activities and provide warning on possible terrorist attacks.

The application of pedestrian detection and surveillance detection requires a system that is able to detect human from images or videos [3]. Computer vision is widely used for the implementation of the system. The basic concept of computer vision can be described as the interpretation of data and object recognition. Initially, a static image or video is loaded into a computer program as an array of complex data. The computer is then programmed to process the data and recognize the object. To develop an object recognition system, the most common way is to extract information from the targeted object in the form of feature set. The feature set is a set of data which is able to represent the object and allow the computer to identify the object. Therefore, the feature extraction algorithm is very crucial in the design of a

robust object recognition system. Normally prior to the feature extraction, the input image will be pre-processed to facilitate the feature extraction process.

An object detection system can be designed to detect any object [4], such as human, vehicles, buildings and animals, which make the object detection a popular field of interest to many researchers worldwide. Human detection is one of the most researched areas, as it is applicable in various applications. For example, pedestrian detection system is implemented in vehicles to detect pedestrians on the road and warns the driver when there is a possibility of collision to prevent accident. Another common application is in security system for intrusion detection. Intruders which enter unauthorized area will be identified as an abnormal event. The system will then send notification to the security personnel for proper action. Such application is also known as abnormal event detection, which is also able to detect human fights, unusual single-individual loitering, vandalism and terrorism activities. Other than that, human detection is also widely used in dense crowds monitoring and people counting. For example, the managements of a shopping mall can count the number of visitors, as well as surveying popular spot for businesses. Another possible application of human detection that may possibly save life is the automatic fall detection system for the care of elderly people.

However, detection of human from images is a challenging task. This is because human always appear in various types of clothing and in different manners of body movement [5]. Moreover, the unstructured and complicated background conditions are another challenge which will cause difficulty in the segmentation and detection. Over the years, researchers have been working on different techniques to improve the human detection accuracy, as well as the time-efficiency of the human detection system to enable it to be applicable in real time.

1.2 Problem Statement

Pedestrian detection targets to detect human that present in a road scene [3]. However, human always appear in different postures and clothing. Various human appearances can cause miss detection. In addition, the possible cluttered background objects on the road such as lamp posts, trees and commercial boards often caused false detection of human. Therefore, extracting features from human in a road scene is a challenging task.

There are a few methods proposed by researchers which present high detection efficiency. However, the complexity of their feature extraction methods caused them to be very computationally extensive and time consuming. Moreover, most of these methods created a feature set with large dimension. These two factors can cause the human detection system to be inefficient and time consuming. The inefficiency of existing human detection method is to time consuming to be implemented in real time.

1.3 Research Objective:

The main objective of this project is:

- **To develop an efficient feature extraction algorithm for human detection based on the histogram of pixels' gradient orientation.**

The research is targeted to find the best parameters setting for the proposed feature extraction method and evaluate the performance of the proposed algorithm and compare it with other existing methods

1.4 Project Scope

In this research, the algorithm development is done by implementing the histogram of oriented gradient, targeting to develop a time-efficient human detection algorithm. However, the problem of occlusion is not being considered in the proposed algorithm.

This research will focus on detecting human during day time using visual camera. During night time, the captured image may be too dark and may not be able to be processed using the proposed method. An infrared camera has to be used to capture night image, which is not within the scope this research.

In addition, this research does not consider the processing speed-up using hardware acceleration methods such as GPU or FPGA. All improvements in the speed performance are based on the development of an efficient feature extraction algorithm. The implementation and evaluation of the proposed method was done on software by using C++ programming. The purpose is to develop a time efficient feature extraction algorithm for human detection and compare it with other algorithms under the same platform.

1.5 Thesis Outline

This thesis is organized into 5 chapters. Chapter 1 provides a brief introduction on vision-based human detection. The objectives and the scopes of the research are also covered in this chapter.

Chapter 2 reviews the current related researches on human detection. Different existing methods of human feature extraction were investigated. In Chapter 3, the methodology of this research is discussed. The proposed feature extraction algorithm will also be explained in detail in this chapter. In Chapter 4, the experiments and results to find the optimal parameters setting for the proposed method is presented.

The evaluation on the performance of the proposed method compared to other existing methods will also be given in this chapter. Finally, the conclusion and some future recommendations for this research are given in Chapter 5.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

In recent years, human detection from images or videos has become a popular field of research in computer vision. The most common approach of human detection is extracting informative features from images and then using a classifier to distinguish between a human and non-human [6].

Feature extraction generates a set of data which represents the image in the form of feature vector. One of the most common features for human detection is the Histogram of Oriented Gradient (HOG) which was proposed by Dalal and Triggs [5]. HOG is a grid descriptor which extracts the feature by constructing histograms of image gradient's orientation in a dense grid. Linear SVM is used in their method to classify the image as human or non-human. However, this HOG feature extraction method creates a large feature pool, which caused the feature extraction and classification process to be very time-consuming. Over the years, many approaches have been proposed by different researchers to speed up the HOG feature extraction.

This chapter is divided into 5 sections. The first section explains the original HOG feature extraction proposed by Dalal and Triggs. In the second section, a literature review on some of the existing approaches done to improve the HOG extraction method is presented. The third section explains the LBP feature extraction, followed by the explanation of the PCA in the fourth section. The fifth section explains the SVM classification and finally a summary of this chapter is given in the last section.

2.2 Feature extraction for human detection

Feature extraction is a method which enables the extraction of important features from a targeted object. For human detection, features that are extracted should contain relevant information of the human figure which is the edges and gradients in different orientations. The existing representative work on feature extraction for human detection includes the Histogram of Oriented Gradient (HOG) [5] and Local Binary Pattern (LBP) [7].

2.2.1 Histogram of Oriented Gradient

The HOG feature extraction method proposed by Dalal and Triggs [5] makes use of gradient orientation and normalized histogram. Figure 2.1 shows the summary of their proposed method. The general ideas of the algorithm comes from calculating the gradient orientation of all of the pixels in the input image and voting the gradients into histograms to form a feature set that can represent the image. Their human detector is divided into two stages, which are the feature extraction phase and the classification phase.

For the feature extraction phase, the input image is initially resized to the size of 64×128 pixels. As the input image has 3 color channels, gamma normalization is done to each of the color channel. Next, a mask will slide through the x and y axes of the image and compute the gradient of each pixel for all the three color channels. The gradient and magnitude are then calculated. After the calculation of the pixels' gradient orientations and magnitudes for the three color channels, the channel with the largest magnitude is selected to be the representative orientation and magnitude for the particular pixel. Its magnitude is then used as a vote to construct the histogram later on based on the gradient orientation.

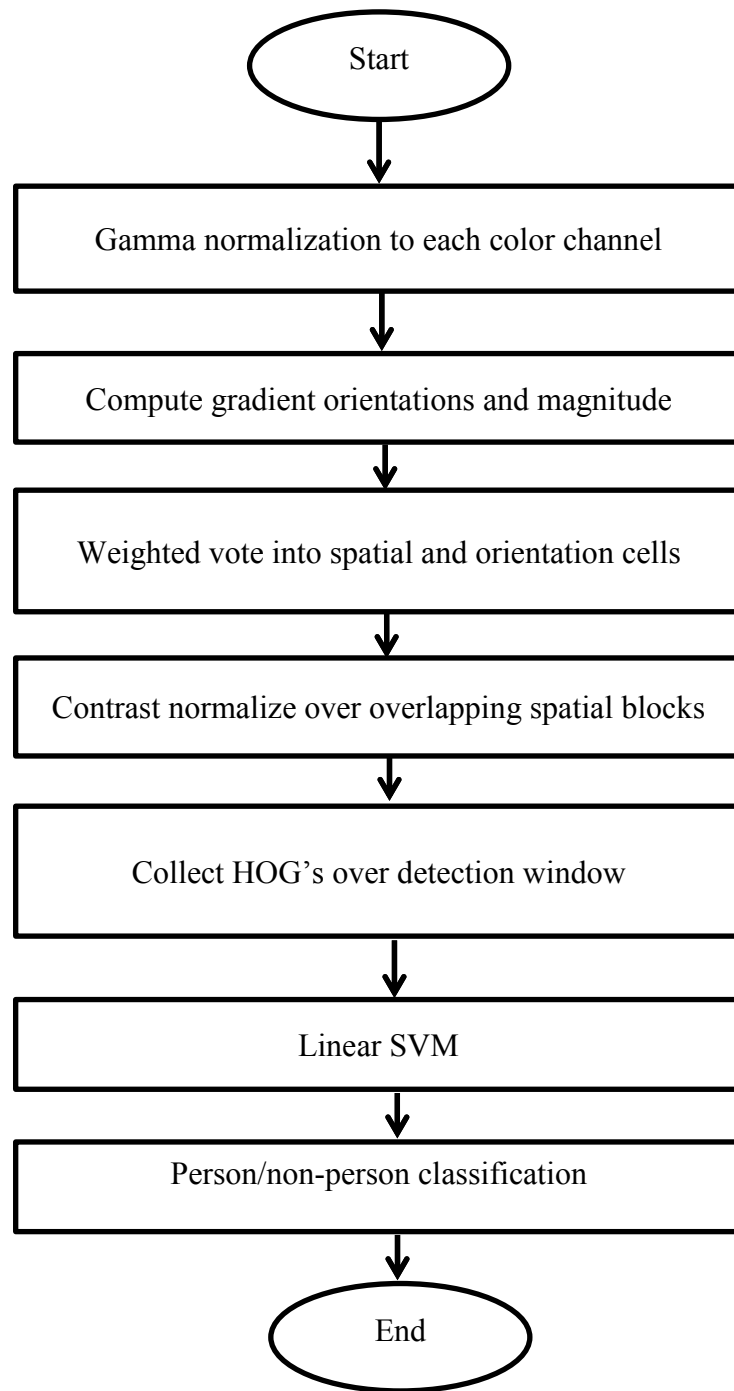


Figure 0.1: Summary of HOG extraction

After the calculation of gradients and magnitude, the resulting image is divided into grids of cells of 8×8 pixels. A 16×16 sliding window will slide through the grid of cells, forming overlapping blocks. Each block consists of four neighboring cells. A histogram is constructed for each cell in the block based on the gradient

magnitude as shown in Figure 2.2. Trilinear interpolation is used to vote the gradient magnitude into the four histograms by referring to the gradient orientation. The formula of trilinear interpolation is defined as:

$$\begin{aligned}
\mathbf{h}(x_1, y_1, z_1) &\leftarrow \mathbf{h}(x_1, y_1, z_1) + w(1 - \frac{x-x_1}{b_x}) (1 - \frac{y-y_1}{b_y}) (1 - \frac{z-z_1}{b_z}) \\
\mathbf{h}(x_1, y_1, z_2) &\leftarrow \mathbf{h}(x_1, y_1, z_2) + w(1 - \frac{x-x_1}{b_x}) (1 - \frac{y-y_1}{b_y}) (\frac{z-z_1}{b_z}) \\
\mathbf{h}(x_1, y_2, z_1) &\leftarrow \mathbf{h}(x_1, y_2, z_1) + w(\frac{x-x_1}{b_x}) (\frac{y-y_1}{b_y}) (1 - \frac{z-z_1}{b_z}) \\
\mathbf{h}(x_2, y_1, z_1) &\leftarrow \mathbf{h}(x_2, y_1, z_1) + w(\frac{x-x_1}{b_x}) (1 - \frac{y-y_1}{b_y}) (1 - \frac{z-z_1}{b_z}) \\
\mathbf{h}(x_1, y_2, z_2) &\leftarrow \mathbf{h}(x_1, y_2, z_2) + w(1 - \frac{x-x_1}{b_x}) (\frac{y-y_1}{b_y}) (\frac{z-z_1}{b_z}) \\
\mathbf{h}(x_2, y_1, z_2) &\leftarrow \mathbf{h}(x_2, y_1, z_2) + w(\frac{x-x_1}{b_x}) (1 - \frac{y-y_1}{b_y}) (\frac{z-z_1}{b_z}) \\
\mathbf{h}(x_2, y_2, z_1) &\leftarrow \mathbf{h}(x_2, y_2, z_1) + w(\frac{x-x_1}{b_x}) (\frac{y-y_1}{b_y}) (1 - \frac{z-z_1}{b_z}) \\
\mathbf{h}(x_2, y_2, z_2) &\leftarrow \mathbf{h}(x_2, y_2, z_2) + w(\frac{x-x_1}{b_x}) (\frac{y-y_1}{b_y}) (\frac{z-z_1}{b_z})
\end{aligned} \tag{2.1}$$

where the w is the weight to be distributed among the neighboring bins, b is the inter-bin distance, x is a point located in between two neighboring bins x_1 and x_2 . When the bilinear interpolation is extended to trilinear interpolation, it has y which is a point located in between y_1 and y_2 , while z is a point between z_1 and z_2 . The weight will be distributed among the nine neighboring bins.

For the voting of orientation angle into the histogram, the orientation angles from 0 to 180° are evenly divided into 9 histogram bins as shown in Figure 2.3. For the angles from 180 to 360° , the orientations angles are flipped by 180° into the range from 0 to 180° and voted into the 9 histogram bins.

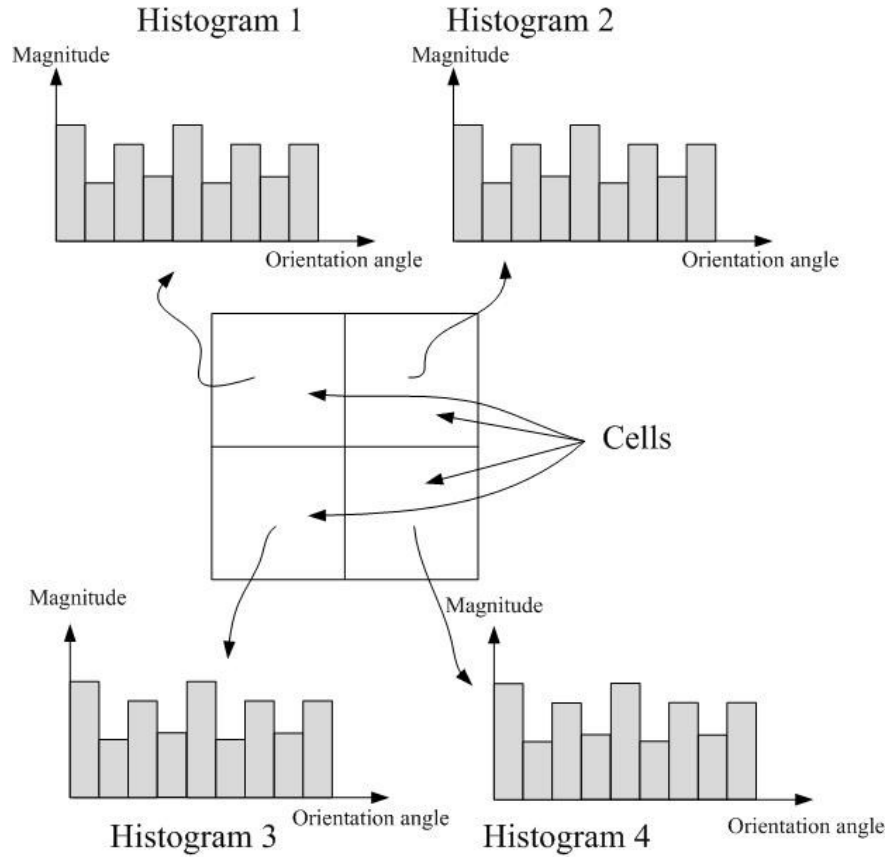


Figure 0.2: Block formed from 4 neighboring cells

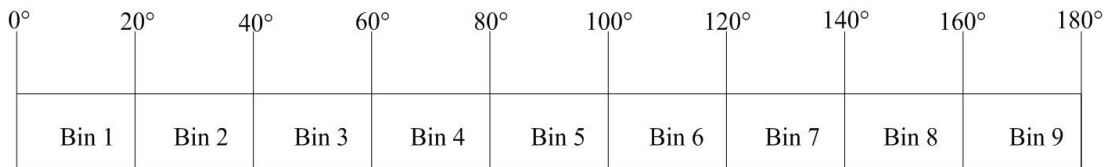


Figure 0.3: The 9 histograms bins and their respective ranges of orientation angles

For a 64×128 pixels input image, a total of $8 \times 16 = 128$ cells can be formed, with each cell consists of 8×8 pixels. As the sliding window slides across the cells, overlapping each other, a total of $128 - 16 - 7 = 105$ blocks can be formed, with each block covering 2×2 cells. Next, normalization is performed on each block. For the final step, the histogram values for each of the 105 blocks are collected and formed a feature vector.